REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
16-05-2014	R&D Status Report (Quarterly)	17-02-2014 to 16-05-2014

4. TITLE AND SUBTITLE
Construction of a Bacterial Cell that Contains Only the Set Of Essential Genes Necessary to Impart Life

5a. CONTRACT NUMBER
HR0011-12-C-0063

5b. GRANT NUMBER
NA

5c. PROGRAM ELEMENT NUMBER
NA

6. AUTHOR(S)
John Glass

Tony Yee

5d. PROJECT NUMBER
NA

5e. TASK NUMBER
NA

5f. WORK UNIT NUMBER
NA

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
J. Craig Venter Institute
9704 Medical Center Drive
Rockville, MD 20850

8. PERFORMING ORGANIZATION REPORT NUMBER
HR0011-12-C-0063.8

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA, MTO
Prog: Living Foundries ATGC
675 N Randolph St
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
DARPA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
NA

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES
NA

14. ABSTRACT

We now have a viable reduced genome design (RGD) cell that has a genome that is 85% minimized. This is a genome in which six and one half of the eight genome segments are minimized. The key issue that we must resolve to produce a fully minimized RGD is that in RGD segments 1 and 3 there are at least two instances of what we call synthetic lethal deletions that we have to identify and add back to the genome.

The top down approach of iteratively removing non-essential genes from M. mycoides syn1.0 continues. Genes or sets of genes are deleted from previous top down constructs while the genomes are parked in yeast cells. Then the reduced genomes are isolated from yeast cells and transplanted to determine if the deletion is viable or has an altered phenotype. Currently we are working on our 17th iteration of this process. The smallest genome the top down effort has made to date has 764 kb and is missing 230 protein coding genes. This deletion mutant grows normally.

The effort to modularize the genome is in progress on multiple fronts.

15. SUBJECT TERMS
NA

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Tony Yee
UU	UU	UU	UU	9	**19b. TELEPHONE NUMBER** *(include area code)* 301 795 7133

J. Craig Venter™
INSTITUTE

Construction of a Bacterial Cell that Contains Only the Set of Essential Genes Necessary to Impart Life

Report Title:	R&D Status Report (Quarterly)
Report Number:	HR0011-12-C-0063.8
Reporting Period:	February 17, 2014 to May 16, 2014
Contract No.:	HR0011-12-C-0063
Performing Organization:	J. Craig Venter Institute 9704 Medical Center Drive Rockville, MD 20850 USA
Author of Report:	John Glass, Anthony Yee
Principal Investigators:	Hamilton Smith, Clyde Hutchison

Abstract

We now have a viable reduced genome design (RGD) cell that has a genome that is 85% minimized. This is a genome in which six and one half of the eight genome segments are minimized. It grows at less than half the rate of wild type *Mycoplasma mycoides*. This organism, called RGD23*45678 contains only segment 1 and half of segment 3 that are wild type. We have another strain, RGD24678 (meaning it has only 3 wild type genome segments) that grows at the type rate of one doubling per hour. The key issue that we must resolve to produce a fully minimized RGD is that in RGD segments 1 and 3 there are at least two instances of what we call synthetic lethal deletions that we have to identify and add back to the genome. A synthetic lethal is the result of deleting both members of a pair of genes that both encode the same essential function. Individually, either member of the pair can be deleted; however a genome with both genes omitted will not support life. Towards that goal, we have pursued 3 methods of identifying the synthetic lethals so that the RGD genome can be redesigned and synthesized. Based on those studies the RGD2 genome has been designed and synthesis is in progress.

The top down approach of iteratively removing non-essential genes from *M. mycoides* syn1.0 continues. Genes or sets of genes are deleted from previous top down constructs while the genomes are parked in yeast cells. Then the reduced genomes are isolated from yeast cells and transplanted to determine if the deletion is viable or has an altered phenotype. This effort is more about directing the bottom up approach by identifying potential synthetic lethals than about eventually reaching a fully minimized genome by iterative gene deletion. Currently we are working on our 17th iteration of this process. The smallest genome the top down effort has made to date has 764 kb and is missing 230 protein coding genes. This deletion mutant grows normally.

Sponsored by
Defense Advanced Research Projects Agency
Microsystems Technology Office (MTO)
Program: Living Foundries: Advanced Tools and Capabilities for Generalizable Platforms (ATCG)
Issued by DARPA/CMO under Contract No: HR0011-12-C-0063

The views expressed are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.
Approved for Release; Distribution is Unlimited

The effort to modularize the genome is in progress on three fronts. A 30 gene tRNA module was previously constructed and inserted into the *M. mycoides* syn1.0 genome. We are now determining if that genome will remain viable as each of the remaining 12 natural tRNA clusters are removed such that the synthetic genes in the module must support the tRNA needs of the cell. We are designing modules for other functional groups such as glycolysis, glycerol metabolism, and amino acyl tRNA synthetases. Finally, in an effort to determine the functions of genes of unknown function that are similar to genes from other bacteria that encode biochemically characterized proteins, we are swapping native uncharacterized genes for characterized genes from *Bacillus subtilis*, that is presented in a constitutive expression module. If the *B. subtilis* gene containing *M. mycoides* mutant is viable than the function of the conserved hypothetical gene is the same as the input *B. subtilis* gene.

Table of Contents:

Section
Summary 2
Introduction 3
Methods, Assumptions and Procedures 3
Results and Discussion 5
Conclusion 8
Planned Activities for the Next Reporting Period 9
Program Financial Status 9

Summary

The goal of the project is to create a cell that contains only the set of genes that are essential for life under ideal laboratory conditions. We are working to minimize *Mycoplasma mycoides* JCVI-syn1.0 (the synthetic version of *Mycoplasma mycoides* subsp *capri*) using two approaches:

- Top Down: remove genes and clusters of genes one (or a few) at a time, proceeding only if the reduced strain is viable, with a reasonable growth rate
    - We previously reported the results of a transposon study was conducted and allowed us to categorize genes as Essential (E), Non-essential (N), or Impaired (I)
        - This categorization scheme has been the basis of most of our subsequent work
    - The *M. mycoides* genome has been reduced to 764 kb using the Tandem Repeat Endonuclease Cleavage (TREC) strategy
- Bottom Up: design a reduced genome based on our best Tn5 gene disruption and deletion data (RGD), and synthesize it
    - Synthesis from oligonucleotides
        - All 1/8th genome molecules have been tested and found to be viable
        - Viable genomes containing multiple RGD segments have been constructed
        - RGD24678 (clone 19) is viable and grows at the same rate as wild type syn1.0
        - RGD23*45678 (clone 59), which has 6.5 (partial 3) RGD cassettes is viable but grows very slowly
        - Various analyses to identify synthetic lethal genes that must be added back to the RGD genome in order to generate a viable cell have identified 30 potential candidates located in segments 1,2,3 and 5. These are being resynthesized to produce the RGD2 genome, which will likely be viable.

tRNA Modularization: The initial modularization experiments are progressing. A 5.3 kb tRNA module containing the 30 tRNA genes plus the necessary promoters and terminators was constructed and sequence verified. The module was inserted into the genome in place of the largest natural cluster of tRNAs and found to be viable. In Q2 we tested the effects of removing the

remaining 12 tRNA clusters from the cell containing the synthetic module. To date most, but not all are viable.

Interspecies modules to characterize unknown genes: Characterized genes from *B. subtilis* were swapped with similar, but not so similar as to be clearly the same, essential genes from *M. mycoides*. The *B. subtilis* genes are in a high-level expression module. When the resulting genomes are viable if confirms the function of the *M. mycoides* gene.

Introduction

The goal of this research project is to build a minimal bacterial cell that contains only the genes that are required for life in ideal laboratory conditions. The pursuit of a minimized cell is critical to the advancement of biology, both as a pathway for understanding the basic requirements for replication and as a chassis for creating an optimized platform for any number of possible applications.

We previously reported that the *Mycoplasma mycoides* JCVI-syn1.0 genome was successfully reduced from 1078 kb to 779 kb; however, while the 779 kb genome was viable, the growth rate was far too slow to allow follow up experiments at an acceptable pace. Using the N, E, I gene categories, the genome has been reduced to 764 kb, but with a normal doubling time.

As reported previously, all 8 of the 1/8th RGD segments tested have proven to be individually viable when in the complementing 7/8th background. We now have several genomes that contain up to 5 RGD segments, and one genome that has 6.5 RGD segments.

A preliminary tRNA module was designed, constructed, and introduced into the *M. mycoides* syn1.0 genome and found to be viable. Synthesis of segments with the natural tRNA loci removed is in progress.

Methods, Assumptions and Procedures

TOP DOWN APPROACH

The plan here was to start with the full size 1078 kb *M. mycoides* JCVI-syn1.0 synthetic genome. We have continued to use the TREC strategy to make iterative deletions in the mycoplasma genome. Targeting the N category genes and clusters is proving to be effective (further discussed in the Results and Discussion section). We have made a series of strains that are progressively reduced with little to no reduction in growth rate.

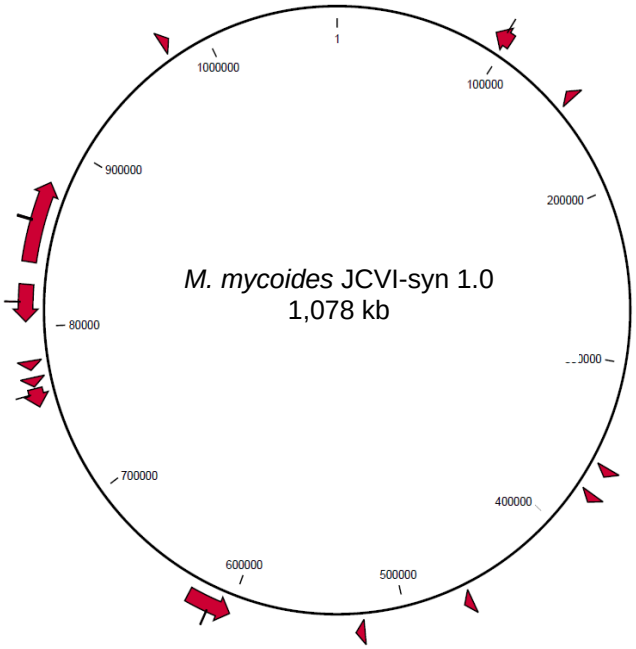
BOTTOM UP APPROACH

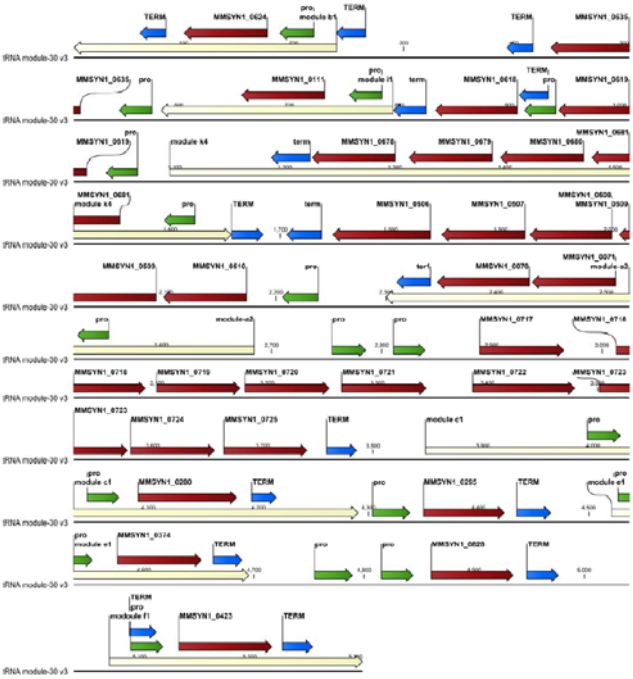
Synthesis from oligonucleotides: A new eight piece genome design was completed using the N, E, I gene classification system (RGD, 539 kb). All 8 have been found to support cellular life. We have continued testing combinations of reduced genome segments. We now have several genomes that contain up to 6.5 RGD segments.

MODULARIZATION

To test gene modularization, we have organized the 30 tRNA genes of *M. mycoides* into a single contiguous module. The module contains the coding regions, as well as the promoters and terminators needed for regulation. The tRNA genes are naturally distributed around the genome in 13 loci.

Figure 1





(a) Natural distribution of tRNA genes in M. mycoides. *The tRNA gene clusters have been enlarged in Fig.1(a) to show the direction of transcription. The M. mycoides JCVI-syn1.0 genome has 8 single tRNA genes and 5 clusters of 2 to 9 genes, for a total of 30.*

(b) tRNA module design. *The 30 tRNA genes have been relocated into a single module. (Green arrows represent promoters. Red arrows show tRNA genes. Blue arrows are terminators.)*

Each of the 13 loci was synthesized by PCR using syn1.0 as the template, cloned in *E. coli* and then joined together into a single cassette with appropriate yeast markers. The cassette was inserted into syn1.0 to replace the largest cluster of 9 tRNAs at 10 o'clock on the genome map. The resulting genome is viable after transplantation. We will synthesize segments with the 12 other tRNA loci around the genome removed from the design.

GENOME COMPLIMENTATION

We have begun initial experiments aimed at enabling genetic complementation to restore desirable phenotypes to deleted strains, for example, RGD. A system capable of quickly adding deleted genes back into a genome would be a powerful tool to help de-convolute growth-retarding synthetic effects.

We are initiating an approach in which we will predict the problematic genes, assemble them into a cassette and insert them into a knock-in strain to test for restored viability.

Results and Discussion

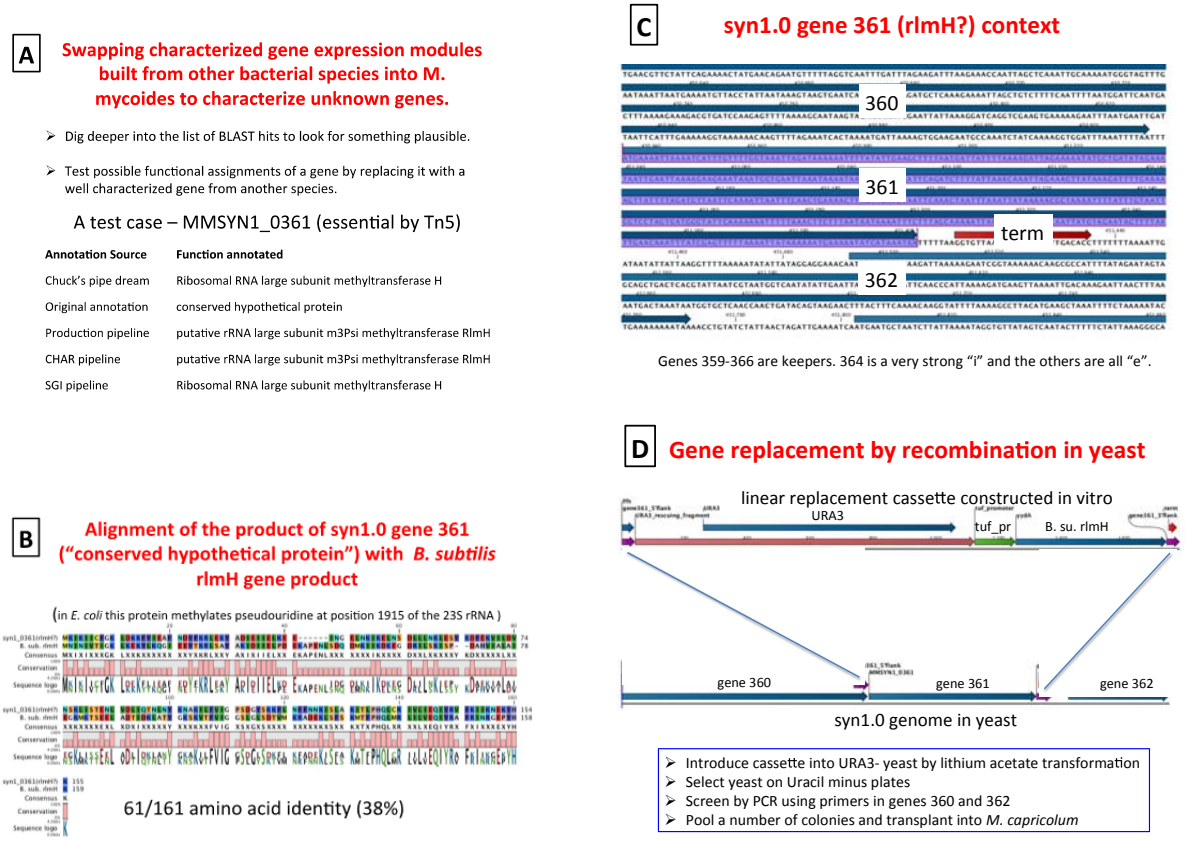
MODULARIZATION

tRNAs: As reported in Q1, we have constructed a synthetic tRNA module that encodes all 30 tRNAs, transcriptional promoters and terminators as a single 5.3 kb cassette. We substituted the cassette for the 9 tRNA cluster of tRNAs in *M. mycoides*, and that cell grew normally. This was an important first step. We now have a cell with a single copy of 9 tRNAs and two copies of the other 22. In Q2 we have tested whether cell made previously would remain viable if individually removed the other 12 clusters of tRNAs. To date 8 of the 12 tRNA cluster deletions yielded viable *M. mycoides*; although we were surprised that some of these grow slower than the wild type cell. Our plan is to repeat efforts to show that the remaining four clusters of tRNAs can be deleted and still result in a viable cell.

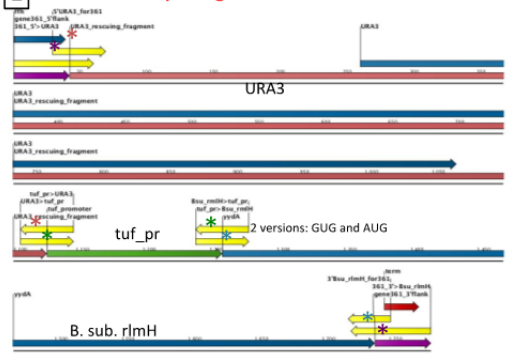
The next phase in this effort will be to build RGD (reduced genome design) minimal cell modules with the tRNA modules replaced by a transcriptional terminator (we do not want unintended transcription from the genes flanking the tRNA site disrupting the cell). As we have already shown to be viable, the 9 tRNA cluster will be replaced with the 30 tRNA module.

Module swap to characterize genes of unknown function: In another aspect of the modularization effort, we have replaced an essential gene annotated in *M. mycoides* as a conserved hypothetical protein with a similar gene expressing a characterized enzyme.

Figure 2



E **Cassette to replace gene 361 with *Bacillus subtilis rlmH***



- It appears we have replaced the essential gene MMSYN1_0361 with Bsu_rlmH and the mycoplasma is fully viable.
- Since the *B. subtilis* gene is known to encode RNA large subunit methyltransferase H, we infer that MMSYN1_0361 has the same function.
- There are a couple of controls I plan to do:
  1) Screen for an internal fragment of 361
  2) Replace 361 by just URA3. This should be dead.
- We should develop a list of candidate genes from the RGD for applying this approach to confirm gene function.

F **Positive clones for each construction were pooled & transplanted**

- Positive yeast clones were pooled and grown up in liquid.
- Plugs were prepared.
- Bogumil transplanted 2 yeast plugs into *M. capricolum*
- All three constructs gave transplants with colonies similar in size to syn1.0

Bsu_rlmH (GUG)	11 colonies
Bsu_rlmH (AUG)	13 colonies
MMSYN1_0361	21 colonies

- Colonies picked into 1 ml SP4+, grown overnight
- 1 ul used in MXPCR to screen junctions

Fig. 2. Swapping characterized gene expression modules built from other bacterial species into *M. mycoides* to characterize unknown genes. (A) In this example of this approach of using synthetic expression modules based on genes of known function in other organisms to determine the function of similar genes in *M. mycoides* but not so similar that they can be definitively annotated we looked at *M. mycoides* essential gene 361. (B) This gene was originally annotated in *M. mycoides* as a conserved hypothetical protein. The encoded protein is 38% identical to a characterized gene in *B. subtilis* that encodes the ribosomal RNA large subunit methyltransferase H. (C) The locale of gene 361 in the *M. mycoides* chromosome. (D) To swap the *B, subtilis* gene for the *M. mycoides* gene in yeast, a cassette containing a URA3 marker and the *B. subtilis rlmH* gene was constructed with the rlmH gene behind a *M. mycoides tuf* promoter. It was exchanged into the YCp and that was transplanted into *M. capricolum*. (E) PCRs were done using primers at the asterisks to confirm that the resulting transplants had the desired genes. (F) Analysis confirmed that *B. subtilis rlmH* could replace the *M. mycoides* 361 gene and that the 361 gene likely encodes ribosomal RNA large subunit methyltransferase H. We envision using this method to evaluate the function of many of the unknown genes in the minimal cell.

TOP DOWN APPROACH

Iterative deletions using the TREC based approach are making steady progress toward a minimal genome. A table outlining the progress to date is shown below. Since the last reporting period, strain D16 has been tested and found to be viable with a good qualitative growth rate (quantitative growth rate evaluation has not been performed).

Table 1

Strains	DT(min)	Genome Size (bp)	# of Genes Deleted
syn1.0	64	1,078,809	0
syn1.0D6 RE		1,062,183	17
DISs		1,048,690	31
D1		979,083	68
D2		969,069	74
D3		944,159	90
D4		931,710	97
D5		923,647	102
D6	67	908,931	108
D7		877,942	135
D8		866,271	155
D9	64	844,265	173
D10	65	828,901	181
D11		816,807	194
D12		805,506	201
D13		794,666	200
D14		784,762	207
D15		775,131	216
D16		763,899	224
D17*		749,766	235

*In progress

Sequential deletion of genes and clusters of genes is not the way that we ultimately expect to pursue a minimal genome design. While we are making steady progress in this approach, the greater outcome of performing the top down deletions is the generation of information regarding unanticipated interactions between elements of the genome that are difficult or impossible to identify using the Bottom Up approach. De-convolution of synthetic lethal effects will be difficult work, which we hope to delay as long as possible with information gained from the top down approach.

BOTTOM UP APPROACH

Work this quarter has centered on identification of the synthetic lethal genes, and design of a new genome that encodes the essential functions. Not only do we want to identify the problematic genes, we also want to evaluate different methods for making the identification. We are going about this process using multiple approaches:

- From both our own analyses and from literature data on related mycoplasmas, we have come to realize that glycerol metabolism is essential and that we had deleted all the genes capable of importing glycerol into the minimal cell. It was the hypothesis of the *M. mycoides* research community that if the glycerol metabolism genes could be deleted, the resulting cell would be

non-pathogenic and possibly usable as a live attenuated vaccine. This is because one of the byproducts of glycerol metabolism in *M. mycoides* is hydrogen peroxide, which is a virulence factor (other subspecies of *M. mycoides* are a major cattle pathogen in sub-Saharan Africa and there is no good vaccine). Our work has dashed that hypothesis. The operon comprised of glycerol kinase, glycerol oxidase and the glycerol uptake facilitator will be put back in RGD segment 3. Alternatively we could have put back glycerophosphodiester phosphodiesterase into segment 2 and a glycerol-3-phosphate transporter protein into segment 5. Similarly, we are re-examining the genes we have removed that have annotations in hopes of realizing based on our knowledge of biology that we have removed genes known to encode essential functions.

- We have done tn5 bombardment of on *M. mycoides* mutants that were 7/8^{ths} wild type and 1/8th RGD to look for genes that were not essential in the fully wild type *M. mycoides*, but could not be hit with transposons when sets of genes had been removed. This pointed to a series of potential problem genes that we are considering putting back in the RGD. We were expecting only one or two genes to become essential in each 7/8th wild type 1/8th RGD; however we found many more than that, which has made analysis more complicated (we did see that glycerol metabolism genes 217-219 became essential when the aforementioned genes in segments 2 or 5 were missing).
- In still another approach using a fast growing RGD24678 clone, we used TREC to one at a time systematically delete all the clusters of genes that we originally thought were not essential in segments 1 and the part of segment 3 that was RGD in RGD23*45678 Clone 59. Once again, the data point to more than a single gene or group of genes to be restored to the RGD. This could be because there are multiple synthetic lethals to be repaired in segments 1 and 3 or because of other problems not related to the synthetic lethals.
- Unexpectedly, we found that cells containing RGD4 could not be transformed with tn5. By doing complementation studies in RGD5 cells we have determined that at least one of the genes in a cluster of seven genes omitted from RGD4 is necessary for tn5 transformation.
- Based on our analyses using the methods described above, we redesigned the RGD to make RGD2, which will contain 26 genes previously omitted from RGD that we have evidence might be needed for viability (and tn5 transformability). We chose these genes to both generate a viable cell and a cell that grows at the same rate as wild type *M. mycoides*. We realize some of these added genes are probably not essential, but the priority is to generate a viable near minimal cell with each 1/8th genome segment almost fully minimized. This RGD2 design, at 574,572 kb will still be smaller than the genome of *Mycoplasma genitalium,* which at 579,508 bp has the smallest genome of any organism that can be grown in axenic culture. Later we can do an additional tn5 bombardment of the RGD2 genome to identify any genes that can still be deleted from the genome either by TREC or re-synthesis.

Conclusions

Tasks from the Statement of Work:

Task 1: **Complete a detailed global Tn5 transposon mutagenesis insertion map.**
The Tn5 transposon insertion map was submitted with the initial quarterly report.

Due: Month 6; Status - complete

Task 2: **Delete up to 27 large gene clusters**
Using top down approaches we have reduced the genome size of *M. mycoides* JCVI-syn1.0 from 1079 kb to 764kb through the deletion of some >35 clusters, representing a ~30% reduction. Using the bottom up approach we have a genome with ~85% of the non-essential genes removed in a ~590 kb genome, which makes it the second smallest cell known to grow I pure culture.

Due: Month 12; Status – complete

Task 3: Construct a preliminary modular map of the genome
The design of a modular map of the genome is complete was presented in the May 2012 quarterly report

Due: Month 12; Status – complete

Task 4: Make new transposon insertion map. Identify non-essential small 2-4 gene clusters. Delete small clusters.
A transposon study was performed and previously reported. A table showing the resulting N/E/I categorization system was presented in May 2013 (Deliverable 1). In Q2 we performed tn5 bombardment on our individual RGD cassettes as well as some of the multiple cassette strains in order to identify synthetic lethals and identify remaining non-essential genes.

Once the RGD is tested and viable, we will use Tn5 mutagenesis to discover if the classifications of any of the remaining genes have changed and to determine and report which additional genes might be removed (Deliverable 2).

Planned Activities for the Next Reporting Period

1. Continue the Top Down minimization of our synthetic genome (near term) and conduct a new transposon study on the minimized genome.
2. Based on synthetic lethal analyses, we will synthesize segments RGD2-1, RGD2-3, RGD2-4, and RGD2-5. These will be assembled with existing segments RGD2, RGD6, RGD7 and RGD8 to make the RGD2 genome. We will test this for viability.
3. Continue with verification and testing of the tRNA gene module.
4. Continue swapping expression modules from characterized *B. subtilis* genes with similar but as yet uncharacterized essential genes from *M. mycoides* to determine the functions of the unknown genes.
5. Write manuscript(s)describing the RGD cell and early efforts at modularization.

Program Financial Status

In Process & Completed Tasks	Planned Expend	Actual Expend (Cumulative to Date)	% Budget Completion	At Completion	Latest Revised Estimate	Remarks
Task 1	$305,646	$305,646	100%	$305,646	$305,646	Completed
Task 2	$826,256	$798,351	97%	$798,351	$826,256	Completed
Task 3	$43,487	$43,487	100%	$43,487	$43,487	Completed
Task 4	$1,100,000	$1,084,923	99%	N/A	$1,100,000	In Progress
Cumulative	$2,275,389	$2,232,407	98%	N/A	$2,275,389	N/A

There is no management reserve or unallocated resources. The financial data presented is current through Dec. 2013.

Based on the currently authorized work:

- Is current funding sufficient for the current fiscal year? Yes
- What is the next fiscal year funding requirement at current anticipated levels? $1,214,151.00
- Have you included in the report narrative any explanation of the above data and are they cross-referenced? Not applicable; current funding is sufficient for the current fiscal year.